AD-A078 227 PRINCETON UNIV NJ DEPT OF STATISTICS F/G 12/1
CONSISTENT ESTIMATION OF VARIANCE PARAMETERS FROM MANY SMALL SA--ETC(U)
OCT 79 N P JEWELL, G M RAAB N00014-79-C-0322
UNCLASSIFIED TR-159-SER-2 NL

1 OF 1
AD
A078227

END
DATE
FILMED
1-80
DDC

LEVEL II

(6)

AD A078227

(14) TR-159-SER-2

DDC
RECEIVED
DEC 17 1979
E

(6) CONSISTENT ESTIMATION OF VARIANCE PARAMETERS FROM MANY SMALL SAMPLES WITH DIFFERENT MEANS

by

(10) Nicholas P. Jewell*
~~Princeton University~~
~~and~~
Gillian M. Raab
Medical Computing and
Statistics Unit
University of Edinburgh

(12) 17

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DDC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Technical Report No. 159, Series 2
Department of Statistics
Princeton University
(11) ~~October~~ Oct 1979

*Research supported in part by a contract with the Office of Naval Research, No. (15) N00014-79-C-0322, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

(9) Technical rept.,

406 873 Gm

SUMMARY

Where sets of observations are normally distributed with variances which are related to the mean of each set, the mean values become nuisance parameters when we wish to pool information about the variances from a large number of sets of observations. This paper considers the problem of obtaining consistent estimators of the variance parameters where no assumptions or prior knowledge are available about the mean values. For the case when the variance is proportional to the square of the mean we obtain an estimator for the constant of proportionality which is always consistent, by a marginal likelihood approach. However, this method cannot be generalized to other variance functions. Integrated likelihoods, modified likelihood and partial conditional likelihood methods are investigated for this example and suggest methods for the general example when the standard deviation is small compared with the mean. The partial conditional likelihood method may be a useful general method of eliminating nuisance parameters in problems where the methods proposed by Kalbfleisch and Sprott (1970) are not applicable.

ACKNOWLEDGEMENTS

The authors would like to acknowledge discussions with A. M. Davie, V. Fidler and A. F. Siegel on some points in this paper.

1. INTRODUCTION

Suppose x_j is a random variable (possibly a vector) with a probability distribution which depends on a set of structural parameters θ and a set of incidental parameters τ. The definitions of structural and incidental parameters were given by Neyman and Scott (1948) where they illustrated some difficulties in using maximum likelihood estimation of structural parameters in the presence of a large number of incidental (nuisance) parameters. Many papers have tackled the problem of eliminating unwanted parameters so that statements may be made about the parameters of interest. Most methods have involved some adaptation of likelihood techniques and as early as 1937 Bartlett (1937) discussed such estimation problems in the presence of nuisance parameters, in particular, in the context of his test for the homogeneity of variance estimates. A useful survey and introduction to these adapted likelihood approaches is given by Kalbfleisch and Sprott (1970). The properties of these adapted likelihood estimates are of interest, in particular whether they are consistent. The nost notable result in this direction is due to Andersen (1970) who proved that the conditional maximum likelihood estimator (see Section 2) is always consistent under suitable regularity conditions.

Raab (1979) considered the problem of obtaining consistent estimates of parameters in variance functions by pooling information from a large number of small samples. This problem has direct application in the field of immunoassays, where for each of a large set of different levels of response one obtains a small number of replicate observations (usually radioactive counts). Here the variance of the responses usually increases with their mean level.

Formally we have N sets where the j^{th} member of the i^{th} set is x_{ij} $(j=1,\ldots,r_i > 1)$ which is normally distributed with mean μ_i and variance

$$var(x_{ij}) = \sigma^2 V(\mu_i, \underset{\sim}{\theta}),$$

where $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)$, σ^2 and $\mu_1,\ldots,\mu_N$ are unknown parameters. Possible forms of the variance function V are

$$\text{(i)} \quad V(\mu,\underset{\sim}{\theta}) = 1 + \theta\mu^2, \text{ or}$$

$$\text{(ii)} \quad V(\mu,\underset{\sim}{\theta}) = |\mu|^{\theta} \quad .$$

The number of sets, N, is large but each of the r_i is small. Here $\underset{\sim}{\theta}$ and σ^2 are considered to be structural parameters, or parameters of interest, while μ_i $(i=1,\ldots,N)$ are incidental or nuisance parameters, since our interest in them is restricted to their contribution to the estimation of the variance function. We are interested in the consistency of estimates as $N \to \infty$ while each r_i is fixed. Rodbard et al. (1976) used regression methods for the case when all the r_i are equal *and Finney and* Phillips (1977) used full maximum likelihood estimation of the $N + m + 1$ parameters, $\mu_1,\ldots,\mu_N$, $\theta_1,\ldots,\theta_m$, σ^2. However, it is well known that when $V \equiv 1$ full maximization of the likelihood leads to an inconsistent estimator for σ^2 as $N \to \infty$. Many approaches have been described for producing a consistent estimate of σ^2 for this simple case. In this paper, we illustrate the problems of extending these methods to the general case when V is a function of μ. We also discuss the method of modified likelihood suggested by Raab (1979) for this problem. We are only able to produce analytical results for the case $\theta = 2$, i.e., $V(\mu) = \mu^2$, but this example is of interest as a test case for discussing various

adaptive likelihood approaches and yet is not as simple as the trivial case $V \equiv 1$. This example is of interest in itself, as it is the model for observations with a constant coefficient of variation as the mean changes. The distribution of measurement errors for certain analytes in clinical chemistry may be a potential application. (Healy, 1979)

Raab's method does not suggest a procedure for the general example of estimation of structural parameters in the presence of a large number of incidental parameters in situations where the methods discussed by Kalbfleisch and Sprott (1970) are not applicable. In Section 5 we discuss a general approach initially suggested by Jewell (1979) in the context of estimating a variance parameter in the problem of straight line fitting with heteroscedastic errors in both variables. We examine this method and study its asymptotic properties for the example above where $V(\mu) = \mu^2$ and compare it with the other methods.

We should note here that if $\mu_1,\ldots,\mu_N$ *are assumed to arise* as outcomes of independently and identically distributed random variables with distribution function G_0 (unknown) rather than being considered as unknown parameters then Kiefer and Wolfowitz (1956) have shown that with suitable regularity conditions and conditions on G_0 the full maximum likelihood estimator of σ^2 is consistent for a given function V. G_0 need not be assumed to belong to a parametric class of distributions. However, there are many practical situations such as those arising in immunoassay where these assumptions are either not satisfied or not realistic.

2. CONDITIONAL AND MARGINAL LIKELIHOODS

Consider the problem stated formally in the introduction with variance function V. The total likelihood is given by

$$L(\mu_1,\ldots,\mu_N, \underset{\sim}{\theta}, \sigma^2) = \prod_{i=1}^{N} L_i$$

where

$$L_i = \{2\pi\,\sigma^2\, V(\mu_i,\underset{\sim}{\theta})\}^{-r_i/2} \exp\{-\sum_{j=1}^{r_i}(x_{ij}-\mu_i)^2/2\sigma^2\, V(\mu_i,\underset{\sim}{\theta})\}. \quad (1)$$

For the simple case $V \equiv 1$ two methods discussed by Kalbfleisch and Sprott (1970) are applicable.

When $V \equiv 1$, the N sample means $m_1,\ldots,m_N$ are jointly sufficient for $\mu_1,\ldots,\mu_N$ when σ^2 is known, and thus we can factorize L as follows:

$$L(x_{ij};\, \mu_1,\ldots,\mu_N, \sigma^2)$$

$$= C_1(x_{ij}, \sigma^2 | m_1,\ldots,m_N)\; C_2(m_1,\ldots,m_N;\, \mu_1,\ldots,\mu_N, \sigma^2).$$

We assume that the factor C_2 which is the likelihood of the N means $m_1,\ldots,m_N$ contains no available information concerning σ^2 in the absence of knowledge of $\mu_1,\ldots,\mu_N$. We restrict attention to C_1 which does not depend on $\mu_1,\ldots,\mu_N$ and maximize it with respect to σ^2 obtaining the consistent estimator

$$\sigma^2 = \sum_{i=1}^{N}\sum_{j=1}^{r_i}(x_{ij}-m_i)^2 \Big/ \sum_{i=1}^{N}(r_i-1). \quad (2)$$

C_1 *is called the conditional likelihood of* σ^2.

Alternatively we can transform the data taking $x_{i1},\ldots,x_{ir_i}$ to $x_{i2}-x_{i1},\ldots,x_{ir_i}-x_{i1}$, x_{i1} for each i and notice that the likelihood (1) factorizes into two distribution functions for each i:

$$L_i(x_{ij};\, \mu_i, \sigma^2) = M_{i1}(x_{ij}-x_{i1}:\, j=2,\ldots,r_i;\, \sigma^2)$$

$$\times M_{i2}(x_{i1}|x_{ij}-x_{i1}:\, j=2,\ldots,r_i;\, \mu_i, \sigma^2).$$

In the belief that the factors M_{i2} contain no available information concerning σ^2 in the absence of knowledge of $\mu_1,\ldots,\mu_N$, we restrict attention to $M_1 = \prod_{i=1}^{N} M_{i1}$ and maximize it with respect to σ^2, again obtaining the consistent estimator (2). M_1 is called the marginal *likelihood of* σ^2.

We now turn to the more complicated case $V(\mu) = \mu^2$. For known σ^2 the minimal sufficient statistic μ_i in sample i is (m_i, s_i^2) where

$$s_i^2 = \frac{1}{r_i}\sum_{j=1}^{r_i}(x_{ij}-m_i)^2 .$$

Conditioning on this statistic leads to the factor C_1 being independent of σ^2 so that the conditional likelihood approach does not work.

We now consider the marginal likelihood approach. We transform *the data as follows:*

$$x_{i1},\ldots,x_{ir_i} \to x_{i2}/x_{i1},\ldots,x_{ir_i}/x_{i1},\, x_{i1} \qquad (1 \le i \le N)$$

and factorize the likelihood into two parts:

(i) the likelihood M_1 of the ratios

$x_{i2}/x_{i1},\ldots,x_{ir_i}/x_{i1}$ (for $i=1,\ldots,N$)

which does not depend on $\mu_1,\ldots,\mu_N$;

(ii) the likelihood M_2 of $x_{11},x_{21},\ldots,x_{N1}$,

given $x_{i2}/x_{i1},\ldots,x_{ir_i}/x_{i1}$ for $i=1,\ldots,N$.

We must ignore any sets where $x_{ij} = 0$ $(j=1,\ldots,r_i)$ which will occur with positive probability only when $\mu_i = 0$, and contain no information on σ^2. Otherwise if $x_{i1} = 0$ we choose another non-zero member of

the i^{th} set as our "first" member. It seems reasonable to assume that M_2 contains no available information concerning σ^2 in the absence of knowledge of $\mu_1,\ldots,\mu_N$. Thus we restrict our attention to the marginal likelihood M_1, which may be written as

$$M_1 = \prod_{i=1}^{N} M_{1i}$$

where M_{1i} is the marginal likelihood of $x_{i2}/x_{i1},\ldots,x_{ir_i}/x_{i1}$, which becomes

$$M_{1i} = \frac{1}{(2\pi\sigma^2)^{r_i/2}} \int_{-\infty}^{\infty} |w|^{r-1} \exp\left\{-\frac{1}{2\sigma^2}\,(a_i w^2 - 2b_i w + r_i)\right\} dw ,$$

where
$$a_i = \sum_{j=1}^{r_i} (x_{ij}/x_{i1})^2$$

$$b_i = \sum_{j=1}^{r_i} (x_{ij}/x_{i1}) .$$

Thus M_{1i} can be calculated explicitly when r_i is odd but involves incomplete gamma functions when r_i is even. For the easiest case $r_i = 3$ (all $i=1,\ldots,N$) we obtain

$$M_{1i} = \frac{1}{2\pi(a_i)^{3/2}} \exp\left\{-\frac{1}{2\sigma^2}\,(r-\frac{b_i^2}{a_i})\right\}(1+\frac{b_i^2}{a_i\sigma^2}) .$$

Maximizing the total log likelihood with respect to σ^2 provides the likelihood equation

$$\sum_{i=1}^{N} (r_i - b_i^2/a_i) = 2\sum_{i=1}^{N} \frac{b_i^2/a_i}{(1+b_i^2/(a_i\hat{\sigma}^2))} \qquad (3)$$

for our estimate $\hat{\sigma}^2$.

Similar (though more complicated) equations for $\hat{\sigma}^2$ can be found for other values of r_i $(i=1,\ldots,N)$ which may be mixed in the sense that r_i need not equal r_j when $i \neq j$. Equations such as (3), which yield our estimates of σ^2 must be solved iteratively in general.

Notice that, in general, equation (3) has N roots none of which may be positive for a given set of data. However, it is easy to see that there can be at most one positive root and this root exists with probability one as $N \to \infty$. The positive root corresponds to the maximum of the marginal likelihood M_1 on the parameter space $(0,\infty)$.

r_i's all equal: When we estimate σ^2 using the marginal likelihood approach we are, in fact, using maximum likelihood to estimate a parameter from a sequence of random variables (namely, $\{x_{12}/x_{11},\ldots,x_{12}/x_{11}\},\{x_{22}/x_{21},\ldots,x_{22}/x_{21}\},\ \ldots$) which have identical and independent distributions which are suitably regular. Standard theory tells us that the likelihood equation has a unique consistent root with probability one as $N \to \infty$; this root corresponds to the unique positive root of the marginal likelihood equation and it is easily seen that this corresponds to the maximum of the marginal likelihood. Thus the maximum marginal likelihood estimate is consistent for σ^2.

<u>r_i's different:</u> To examine consistency in this situation we require that the number of sets corresponding to a fixed value of r_i tend to infinity for <u>each</u> of a finite set of values of r_i and the proportion of each member of this finite set tends to a fixed number as $N \to \infty$. From this point on we refer to this situation as <u>consistently mixed values of r_i</u> (as $N \to \infty$). The comments above apply to each group of sets corresponding to a given r_i and it is easy to see that the 'separate' estimates of σ^2 combine together to provide a consistent estimate of σ^2.

In fact consistency for the case $r_i = 3$ $(i=1,\ldots,N)$ can be verified directly from (3) by a method which illustrates the technique we will use in the following sections.

We can write equation (3) as

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{2m_i^2}{s_i^2 + m_i^2 + 3m_i^2/\hat{\sigma}^2} + \frac{m_i^2}{s_i^2 + m_i^2}\right| = 1 .$$

Each m_i follows a $N(\mu_i, \mu_i^2\,\sigma^2/r_i)$ distribution and, independently, $3s_i^2/\mu_i^2\,\sigma^2$ has a chi-square distribution with 2 degrees of freedom. The quantity

$$T_i(\hat{\sigma}^2) = \frac{2m_i^2}{s_i^2 + m_i^2 + 3m_i^2/\hat{\sigma}^2} + \frac{m_i^2}{s_i^2 + m_i^2}$$

has a distribution which is independent of μ_i, and hence the same for all i. Its expectation exists and so, by the strong law of large numbers

$$\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} T_i\right] = E(T_i) \text{ , for any fixed value of } \hat{\sigma}^2.$$

Now we can evaluate $E(T_i)$ as

$$E(T_i) = \int_{-\infty}^{\infty}\int_{0}^{\infty} T_i(2\pi\,\sigma^2/3)^{-1/2}\exp\left\{-\frac{3}{2\sigma^2}\left(\frac{m_i}{\mu_i}-1\right)^2\right\}\exp\left(-\frac{s_i^2}{2\mu_i^2\sigma^2}\right)d\left(\frac{m_i}{\mu_i}\right)d\left(\frac{3s_i^2}{\mu_i^2\sigma^2}\right)$$

where σ^2 is the true parameter value. This reduces to

$$E(T_i) = \int_{1}^{\infty}\sigma^{-3}\left(\frac{2\hat{\sigma}^2}{z\hat{\sigma}^2 + 3} + \frac{1}{z}\right)\left(\frac{\sigma^3}{2z^{3/2}} + \frac{3\sigma}{2z^{5/2}}\right)\exp\left\{-\frac{3}{2\sigma^2}\left(1-\frac{1}{z}\right)\right\}dz$$

$$= 1 \text{ when } \hat{\sigma}^2 = \sigma^2 .$$

Thus the equation $E(T_i(\hat{\sigma}^2)) = 1$ has a unique solution given by $\hat{\sigma}^2 = \sigma^2$. It is easy to see then that some root of the equation (3) converges in probability to σ^2 as $N \to \infty$. By our earlier comments this root corresponds to the global maximum of the marginal likelihood with probability one as $N \to \infty$.

In summary, the marginal likelihood approach produces a consistent estimator in this special case. However we cannot generalize it to other variance functions since, even for other simple cases (e.g., $V=\mu$, $V=\mu^3$), it is impossible to find the transformation necessary to derive a marginal likelihood which is independent of μ.

3. <u>INTEGRATED LIKELIHOODS AND PARTIAL BAYES TECHNIQUES</u>

One method of eliminating the nuisance parameters $\mu_1,\ldots,\mu_N$ is to combine the likelihood with a prior density for $\mu_1,\ldots,\mu_N$ of the form $p(\mu_1,\ldots,\mu_N;\, \underset{\sim}{\theta},\sigma^2)$ and then integrate to remove the

nuisance parameters. This gives the integrated likelihood

$$IL(\underset{\sim}{\theta},\sigma^2) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(\mu_1,\ldots,\mu_N,\ \underset{\sim}{\theta},\sigma^2)p(\mu_1,\ldots,\mu_N;\ \underset{\sim}{\theta},\sigma^2)d\mu_1,\ldots,d\mu_N. \quad (4)$$

This is then maximized with respect to $\underset{\sim}{\theta}$ and σ^2 to produce estimates.

This method depends on precise knowledge of the prior p which may not be available. If we try to resolve this difficulty by the use of an improper prior for the $\mu_1,\ldots,\mu_N$, our results will depend on the metric $\{\phi(\mu_j,\ j=1,\ldots,N)\}$ in which the prior is locally uniform. For the simple case $V \equiv 1$ taking a prior which is locally uniform in μ_j, i.e., $(\phi(\mu_j) = \mu_j,\ j=1,\ldots,N)$, and maximizing the resulting integrated likelihood produces the consistent estimator (2).

However the case $V(\mu) \equiv \mu^2$ demonstrates that this is not always the best choice of metric. With a prior locally uniform in μ_i we obtain the integrated likelihood of the i^{th} set from (4) as

$$IL_i = \int_{-\infty}^{\infty}(2\pi\ \sigma^2\ \mu_i^2)^{-r_i/2}\ \exp\left\{-\sum_{j=1}^{r_i}(x_{ij}-\mu_i)^2/2\sigma^2\ \mu_i^2\right\}\ d\mu_i$$

$$= (2\pi\ \sigma^2)^{-r_i/2}\exp\left\{-\frac{1}{2\sigma^2}(r_i-d_i^2/c_i)\right\}\int_{-\infty}^{\infty}\left\{(z+d_i/c_i)^2\right\}^{(r_i/2)-1}\exp\left(\frac{-c_i}{2\sigma^2}z^2\right)dz$$

where $\quad c_i = \sum_{j=1}^{r_i} x_{ij}^2$ and $d_i = \sum_{j=1}^{r_i} x_{ij}$.

IL_i can be calculated easily for even r_i (for odd r_i the integral reduces to an incomplete gamma function). For the sake of simplicity consider the case $r_i = 2$ for $i=1,\ldots,N$, which gives

$$IL_i = (2\pi\ \sigma^2)^{-\frac{1}{2}}(c_i)^{-\frac{1}{2}}\ \exp\left\{-\frac{1}{2\sigma^2}\ (2-d_i^2/c_i)\right\}\ .$$

Maximizing the total integrated likelihood $IL = \prod_{i=1}^{N} IL_i$ with respect to σ^2 produces the estimate

$$\hat{\sigma}^2 = 2 - \frac{1}{N}\sum_{i=1}^{N}(d_i^2/c_i)\ . \quad (5)$$

This estimate is bounded by 2 <u>whatever</u> the true value of the parameter σ^2.

For the example $V \equiv \mu^2$ a choice of improper prior for μ_j $(j=1,\ldots,N)$ proportional to $1/|\mu_j|$, uniform in the metric $\phi(\mu_j) = \log|\mu_j|$, leads to the same consistent estimate of σ^2 as we obtained in Section 2.

The integrated likelihood $IL = \sum_{i=1}^{N} IL_i$ where

$$IL_i = \int_{-\infty}^{\infty}(2\pi\ \sigma^2\ \mu_i^2)^{-r_i/2}\ \exp\left\{-\sum_{j=1}^{r_i}(x_{ij}-\mu_i)^2/(2\sigma^2\ \mu_i^2)\right\}\frac{d\mu_i}{|\mu_i|}$$

$$= (2\pi\ \sigma^2)^{-r_i/2}\int_{-\infty}^{\infty}|w|^{(r_i-1)}\ \exp\left\{-\frac{1}{2\sigma^2}(c_i w^2-2d_i w+r_i)\right\}dw\ .$$

This is easily seen to be $|x_1|^{-r_1} M_{1i}$, where M_{1i} is the marginal likelihood considered before. Hence the estimate $\hat{\sigma}^2$ obtained by maximizing IL with respect to σ^2 is identical to the estimate obtained from the marginal likelihood approach and is consistent either when all the r_i's are equal or when the r_i's are consistently mixed as $N \to \infty$.

The use of non-informative priors to represent ignorance about a parameter can produce theoretical difficulties in Bayesian analysis (Dawid et al, 1973), and the consistency of the estimator (5) shows

that their consequences can be serious. There is no formal procedure for choosing non-informative priors which will avoid these difficulties in the general case. One such recommendation is that the metric in which the non-informative prior is locally uniform should be chosen to make the likelihood approximately data-translated (see Box and Tiao, 1973, Chap. 1), although Dawid et al (1973) and the example discussed by Cox (1973) show that this rule can fail. In the case $V(\mu) = |\mu|^\theta$, a choice of prior $1/|\mu|^{\theta/2}$ for μ would satisfy this rule, at least when $\sigma^2 \mu^{\theta-2} << 1$. For $V \equiv 1$ and $V \equiv \mu^2$ we have shown that this leads to a consistent estimator of σ^2. In this case we are unable to obtain an explicit expression for

$$IL_i = \int_{-\infty}^{\infty} (2\pi\sigma^2 \mu_i^\theta)^{-r_i/2} \exp\left\{-\sum_{j=1}^{r_i} (x_{ij} - \mu_i)^2/2\sigma^2 \mu_i^\theta\right\} \frac{d\mu}{|\mu|^{\theta/2}}$$

and evaluation of the total integrated likelihood numerically involves the product of N different numerical integrations which is not practical.

Finally, returning to the case $V \equiv \mu^2$, we show that the estimator (5), obtained from the integrated likelihood with prior locally uniform in μ, is consistent to order σ^4 when $\sigma^2 << 1$, and each r_i is even.

We can write the estimator (5) for $r_i = 2$ as

$$\hat{\sigma}^2 = 2 - 2\sum_{i=1}^{N} \frac{m_i^2}{s_i^2 + m_i^2} \quad .$$

Now writing $A_i = m_i^2/(s_i^2 + m_i^2)$ we can see that the distribution of A_i is independent of μ_i and hence of i. Thus we can drop the suffix i. We shall consider the expectation of A for general r

and $\sigma^2 << 1$ as the result will be used in the following section. We can write

$$A = \frac{rX^2}{\sigma^2 Y + rX^2}$$

where X follows a $N(1, \sigma^2/r)$ distribution and Y follows a $\chi^2_{(r-1)}$ distribution independent of X. The integral expression for the expectation of A is difficult to evaluate so we will use a Taylor expansion to obtain an approximation for small values of σ^2. We can write

$$E(A) = E\{G(X, Y)\}$$

$$= G + \frac{1}{2!}\left\{\frac{\partial^2 G}{\partial X^2} M_2(X) + \frac{\partial^2 G}{\partial Y^2} M_2(Y)\right\} \tag{6}$$

$$+ \frac{1}{3!}\left\{\frac{\partial^3 G}{\partial X^3} M_3(X) + \frac{\partial^3 G}{\partial Y^3} M_3(Y)\right\} + \ldots$$

where G and its derivatives on the right hand side are evaluated at the expected values of X and Y and the M_i's are the central moments.

Note that

$$\begin{aligned} E(X) &= 1 & E(Y) &= (r-1) \\ M_2(X) &= \sigma^2/r & M_2(Y) &= 2(r-1) \\ M_3(X) &= 0 & M_3(Y) &= 8(r-1) \\ M_4(X) &= 3(\sigma^2/r)^2 & M_4(Y) &= 48(r-1) + 12(r-1)^2 \\ &\vdots & &\vdots \end{aligned}$$

and

$$\frac{\partial^2 G}{\partial X^2} = \frac{r\sigma^2 Y(\sigma^2 Y - 3rX^2)}{2(\sigma^2 Y + rX^2)^3}, \quad \frac{\partial^2 G}{\partial Y^2} = \frac{r\sigma^4 X^2}{2(\sigma^2 Y + rX^2)^3} \; .$$

The central moments of X are in increasing powers of σ^2, as are the derivatives of G with respect to Y. Thus we can substitute in (6) to obtain

$$E(A) = 1 - \sigma^2 \frac{(r-1)}{r} + \sigma^4 \frac{(r-1)(r-2)}{r^2} + O(\sigma^6) \; . \tag{7}$$

For $r = 2$ we get $E(A) = 1 - \sigma^2/2 + O(\sigma^6)$, and thus the estimator (6) converges in probability to $\sigma^2 + O(\sigma^6)$ as $N \to \infty$.

A similar but more complicated calculation shows that the estimator obtained by integrating the likelihood with a prior uniform in μ_j $(j=1,\ldots,N)$ for the case $r_i = r$ $(i=1,\ldots,N)$, with r even, converges in probability to

$$\sigma^2 + \frac{2(r-2)}{r(r-1)}\sigma^4 + O(\sigma^6) .$$

It is interesting to consider the shape of the likelihood L_i for the i^{th} set of observations as a function of μ_i for fixed σ^2. Except for the case $m_i = 0$, this likelihood is bimodal with one maximum for $\mu_i > 0$ and one for $\mu_i < 0$, and a zero likelihood at $\mu_i = 0$. The global maximum is that where μ_i has the same sign as m_i. Now if we examine the behavior of this likelihood for small σ^2 we find that it is dominated by the higher of two peaks, and the area under the lower peak decreases with σ^2. Thus the estimator from the integrated likelihood approaches consistency as the likelihood becomes effectively unimodal.

4. MAXIMUM AND MODIFIED LIKELIHOOD FOR THE CASE $V(\mu) = \mu^2$

In this section we examine the behavior of estimates of σ^2 obtained from maximum likelihood and maximizing the modified likelihood introduced by Raab (1979) for the case $V(\mu) = \mu^2$. We believe that this case should illustrate some of the properties of these methods when applied to more complicated variance functions.

We can write the total likelihood as follows:

$$L = \prod_{i=1}^{N} L_i$$

where

$$L_i = (2\pi\,\sigma^2\,\mu_i^2)^{-r_i/2} \exp\left\{-\sum_{j=1}^{r_i}(x_{ij}-\mu_i)^2/(2\sigma^2\mu_i^2)\right\} . \qquad (8)$$

Raab (1979) suggested working with a modified likelihood for this problem which she defined as:

$$Q = \prod_{i=1}^{N} Q_i$$

where

$$Q_i = (2\pi\,\sigma^2\,\mu_i^2)^{-(r_i-1)/2} \exp\left\{-\sum_{j=1}^{r_i}(x_{ij}-\mu_i)^2/2\sigma^2\,\mu_i^2\right\} . \qquad (9)$$

We first consider the full maximum likelihood estimates of σ^2 and $\mu_1,\ldots,\mu_N$. Differentiating log L (taken from (8)) with respect to each μ_i in turn yields the following equations for the estimates $M_{11},\ldots,M_{1N}$ of $\mu_i,\ldots,\mu_N$ and V_1 of σ^2:

$$V_1 M_{1i}^2 = s_i^2 - M_{1i}\,m_i + m_i^2 \qquad (i=1,\ldots,N) . \qquad (10)$$

The maximum likelihood estimate of μ_i is thus

$$M_{1i} = m_i\left(\sqrt{1+4V_1\frac{(s_i^2+m_i^2)}{m_i^2}} - 1\right)\Big/ 2V_1 , \qquad (11)$$

which recognizes that we require the solution of (10) with the same sign as m_i for a global maximum.

Differentiating log L with respect to σ^2 yields

$$V_1 \sum_{i=1}^{N} r_i = \sum_{i=1}^{N} \frac{r_i}{M_{1i}^2}\left\{s_i^2 + (M_{1i} - m_i)^2\right\} .$$

and substituting the expression for s_i^2 from (10) gives

$$\sum_{i=1}^{N} \frac{r_i m_i}{M_{1i}} = \sum_{i=1}^{N} r_i \quad . \tag{12}$$

Although, at first sight, this equation does not appear to involve V_1 it is implicitly contained in M_{1i} which here is a function of V_1 and the observations.

Identical calculations using the modified likelihood Q given in (9) yield the following equation for V_2, Raab's modified likelihood estimate of σ^2:

$$\sum_{i=1}^{N} \frac{r_i m_i}{M_{2i}} = \sum_{i=1}^{N} r_i \; , \tag{13}$$

where, for each i, M_{2i} is the solution of the quadratic

$$\frac{(r_i-1)}{r_i} V_2 M_{2i} = s_i^2 - M_{2i}\, m_i + m_i^2 \quad . \tag{14}$$

It is easily seen that when $r_i = r$ $(i=1,\ldots,N)$ $V_1 = \{(r-1)/r\}V_2$. In both cases for a given r, m_i/μ_i follows a $N(1, \sigma^2/r)$ distribution and $(rs_i^2)/(\mu_i^2\sigma^2)$ follows a chi-square distribution with $(r-1)$ degrees of freedom. From the solutions to (10) and (14) it follows that m_i/M_{1i} and m_i/M_{2i} can both be written in terms of m_i/μ_i and s_i^2/μ_i^2 and thus have distributions which are independent of μ_i (and thus of i).

When $r_i = r$ $(i=1,\ldots,N)$ we can write (12) and (13) as

$$\frac{1}{N}\sum_{i=1}^{N} \frac{m_i}{M_{1i}} = 1 \qquad \text{and} \qquad \frac{1}{N}\sum_{i=1}^{N} \frac{m_i}{M_{2i}} = 1 \; .$$

As $N \to \infty$ the left hand sides of these equations converge in probability to $E(m_i/M_{1i})$ and $E(m_i/M_{2i})$, respectively. We examine the 'limit' equations

$$E\{m_i/M_{1i}(V_1)\} = 1 \tag{15}$$

for the maximum likelihood estimate, and

$$E\{m_i/M_{2i}(V_2)\} = 1 \tag{16}$$

for Raab's maximum modified likelihood estimate. It is easy to see that equations (15) and (16) have unique positive roots which are, respectively, the probability limits of a sequence of roots of the equations (12) and (13) as $N \to \infty$. There is a similar difficulty here to the one encountered in Section 2 when maximizing the marginal likelihood, namely, that equations (12) and (13) have multiple roots in general. However, these equations always have unique positive roots which correspond to the global maxima of the full likelihood and modified likelihood, respectively. Hence the estimate V_1 which arises from maximizing the full likelihood converges in probability to the root of equation (15). Similarly, the estimate V_2 which maximizes the modified likelihood converges to the root of equation (16). When the r_i's are not all equal but are consistently mixed as $N \to \infty$ it is easily seen that the same comments are true.

Neither of the equations, (15), (16), has σ^2 as the root implying that neither of the estimates are consistent. The behavior of the equations is quite different for large and small values of σ^2.

We first consider the important practical situation of small σ^2 $(\sigma^2 \to 0)$. We start with the estimator V_1 which converges in probability to the solution γ of

$$E\{m_i/M_{1i}(\gamma)\} = 1 \qquad \text{as} \quad N \to \infty \; .$$

Substituting the expression (11) for M_{11} this becomes

$$E\left\{\frac{m_i^2}{2(s_i^2 + m_i^2)}\left[1 + \sqrt{1 + 4\gamma(s_i^2 + m_i^2)/m_i^2}\,\right]\right\} = 1 \,. \qquad (17)$$

Now we have evaluated the expectation of $A = m_i^2/(s_i^2 + m_i^2)$ for small σ^2 in Section 3. We can similarly evaluate the expectation given in (17), i.e., $E\left\{(A/2) + \sqrt{((A^2/4) + A\gamma)}\right\}$ although the algebra is a little messy. As noted above, it is easily seen that (17) is not satisfied with $\gamma \equiv \sigma^2$ and the maximum likelihood estimator is not consistent. To discover the extent of this inconsistency for small σ^2 we seek a solution of the form $\gamma = a\sigma^2 + b\sigma^4 + \ldots$. Substituting this into $E\left\{(A/2) + \sqrt{((A^2/4) + A\gamma)}\right\}$ yields

$$E\left\{(A/2) + \sqrt{((A^2/4) + A\gamma)}\right\}$$

$$= 1 - \frac{(r-1)}{r}\sigma^2 + \frac{(r-1)(r-2)}{r^2}\sigma^4 + \gamma - \gamma^2 + O(\sigma^6)$$

$$= 1 - \frac{(r-1)}{r}\sigma^2 + \frac{(r-1)(r-2)}{r}\sigma^4 + a\sigma^2 + b\sigma^4 - a^2\sigma^4 + O(\sigma^6) \;;$$

then, equating coefficients of powers of σ^2 gives

$$a = (r-1)/r \qquad b = (r-1)/r^2 \,.$$

Thus as $\sigma^2 \to 0$, the estimator V_1 converges in probability to

$$\gamma = \frac{(r-1)}{r}\sigma^2 + \frac{(r-1)}{r^2}\sigma^4 + O(\sigma^6) \,.$$

Using the fact that $V_2 = \frac{r}{r-1}V_1$ we can easily deduce the properties of Raab's modified likelihood estimator. It is also inconsistent and, for small σ^2, V_2 converges in probability to $\delta = \sigma^2 + \sigma^4/r + O(\sigma^6)$. Thus for small σ^2 the bias of V_2 is of order σ^4, while that of V_1 is of order σ^2.

We now briefly consider the estimators as $\sigma^2 \to \infty$ which is of theoretical interest rather than practical importance. The probability limit of the estimator V_1 given by γ satisfies

$$1 = E\left\{\tfrac{1}{2}A\left(1 + \sqrt{1 + 4\gamma/A}\right)\right\} \,.$$

Now we can write A as $A = U^2/(V + U^2)$, where U is $N(\sqrt{r}/\sigma, 1)$ and independently V is $\chi^2_{(r-1)}$. As $\sigma^2 \to \infty$ the distribution of U approaches $N(0,1)$ and that of $U^2 = Z$ approaches χ^2_1. Thus the equation for γ becomes:

$$1 = \int_0^\infty\int_0^\infty \frac{z}{2(V+z)}\left\{1 + \sqrt{1+4\gamma(V+z)/z}\right\} \frac{z^{-(1/2)}V^{(r-3)/2}\exp\left\{-\frac{1}{2}(V+z)\right\}}{2^{r/2}\,\Gamma(1/2)\,\Gamma((r-1)/2)}\,dz\,dV.$$

Changing variables to $x = \left\{-\sqrt{z} + \sqrt{z+4\gamma(V+z)}\right\}/2\gamma$ and $y = \sqrt{z}/x$ we obtain

$$1 = \int_0^{(1/2)(1+\sqrt{1+4\gamma})} \frac{y(2\gamma+y)(\gamma+y-y^2)^{(r-3)/2}\,(\gamma+y)^{-(r/2)}\,\Gamma(r/2)}{\Gamma(1/2)\,\Gamma((r-1)/2)}\,dy \,.$$

and a further change of variable to $w = y^2/(\gamma+y)$ gives

$$\frac{\Gamma(r/2)}{\Gamma(1/2)\,\Gamma((r-1)/2)}\int_0^1 \left(\frac{1-w}{2}\right)^{(r-3)/2}\sqrt{w+4\gamma}\,dw - 1 + (1/2r) = 0 \,. \qquad (18)$$

It is easy to see that this equation has a finite root since the left hand side has the negative value $1/r - 1$ when $\gamma = 0$, but tends to $+\infty$ as $\gamma \to \infty$. The roots of (18) can be determined for small r:

r	equation	root
2	$(4\gamma)^{(1/2)}/2\pi + \gamma/2 + (1+4\gamma)\arcsin((1-4\gamma)/(1+4\gamma))/4\pi - 5/8 = 0$	1.2225
3	$(1+4\gamma)^{(3/2)} - (4\gamma)^{(3/2)} - 5 = 0$	2.6525

etc.

For all r's equal Raab's modified likelihood estimator is $r/(r-1)$ times the maximum likelihood estimator. Thus the probability

limit of V_2 as $N \to \infty$ also tends to a finite number as $\sigma^2 \to \infty$ given by the root γ of

$$\frac{\Gamma(r/2)}{\Gamma(1/2)\,\Gamma(r-1)/2} \int_0^1 \left(\frac{1-w}{2}\right)^{(r-3)/2} \sqrt{w+4\frac{(r-1)}{r}\gamma}\; dw - 1 + (1/2r) = 0. \quad (19)$$

For $r = 2$ $\quad \gamma = 2.4450$
$r = 3$ $\quad \gamma = 3.9788$
etc.

In summary, as $\sigma^2 \to \infty$ the probability limits of the two estimators considered here both tend to a finite number and hence both estimates are bounded with probability one as $\sigma^2 \to \infty$. From other considerations it follows that the value of this finite number will tend to infinity as $r \to \infty$. Clearly it is completely inappropriate to use these estimators when σ^2 is likely to be larger than the roots of (18) or (19) for the values of r occurring in the sets of observations. Notice also that this 'boundedness' property of the estimates as $\sigma^2 \to \infty$ was also apparent in the estimate $\hat{\sigma}^2$ given in (5) in Section 3.

5. PARTIAL CONDITIONAL LIKELIHOOD ESTIMATES

In order to use the method of conditional likelihood as described in Section 2, we require a minimal sufficient statistic for the nuisance parameters. In our example with $V(\mu) = \mu^2$ the method breaks down because the likelihood conditional on the observed value of the minimal sufficient statistic is independent of the parameters of interest.

The method of partial conditional likelihood suggested by Jewell (1979) avoids this difficulty. Let $x_1, x_2, \ldots$ be an infinite sequence of vector-valued random variables. Let the distribution

of x_i depend on two parameters θ_i and τ_i and further suppose θ_i is a function of τ_i and another parameter β which is the same for all i. We assume that the probability distribution of x_i possesses a density $f(x_i|\theta_i,\tau_i)$ with respect to some σ-finite measure. We now assume that for each $i=1,2,\ldots$ there exists a minimal sufficient statistic $T_i(x_i)$ for τ_i where for the time being we 'forget' that θ_i depends on τ_i; i.e., T_i is minimal sufficient for that 'part' of τ_i which is separated from θ_i. We presume that the conditional distribution of $x_1,\ldots,x_n$ given $T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n$ still depends on θ_i (and hence on β and $\tau_1,\ldots,\tau_n$) and we assume that T_i does not depend on β. The <u>partial conditional likelihood</u> is then given by

$$\Phi(x_1,\ldots,x_n;\, \theta_1,\ldots,\theta_n | t_1,\ldots,t_n) = \Phi(x_1,\ldots,x_n;\, \beta,\tau_1,\ldots,\tau_n | t_1,\ldots,t_n).$$

This is different from the examples of approximate conditional likelihoods discussed by Sprott (1973) where the conditional likelihood does not depend on the nuisance parameters, but the factor in the likelihood for the distribution of the T_i contains information about β. In our case the partial conditional likelihood contains some of the information about the nuisance parameters.

Now if we maximize the full likelihood with respect to $\tau_1,\ldots,\tau_n$ we obtain the values $\tau_1^0,\ldots,\tau_n^0$ which are in general functions of β. If we now substitute τ_i^0 for τ_i in the partial conditional likelihood, we obtain a likelihood Φ^0 which depends only on the observations and β. Maximizing this expression with respect to β gives the partial conditional maximum likelihood estimator of β. The method can be adapted for cases where T_i depends on β by using the ideas of Kalbfleisch and Sprott (1970). The details will appear elsewhere.

For the case $V = \mu^2$ we can write the full likelihood as

$$L = \prod_{i=1}^{N} (2\pi W_i)^{-r_i/2} \exp\left[-\frac{1}{2W_i}\left\{\sum_{j=1}^{r_i}(x_{ij}-m_i)^2 + r_i(\mu_i - m_i)^2\right\}\right]$$

where $\theta_i = W_i(\sigma^2, \mu_i) = \sigma^2 \mu_i^2$ and $\tau_i = \mu_i$. The minimal sufficient statistic for μ_i is m_i (when we 'forget' that W_i depends on μ_i). The partial conditional likelihood for the observed W_i's is then

$$\Phi = \prod_{i=1}^{N} (2\pi\, \sigma^2 \mu_i^2)^{-(r_i-1)/2} r_i^{-(1/2)} \exp\left[-\frac{1}{2\sigma^2 \mu_i^2}\left\{\sum_{j=1}^{r_i}(x_{ij}-m_i)^2\right\}\right] .$$

Now the maximum of the full likelihood is attained at $\mu_1^0, \ldots, \mu_n^0$ which are given as functions of σ^2 by equation (11) (replacing V_1 by σ^2). Substituting these values in the partial conditional likelihood we obtain

$$\Phi^0 = \prod_{i=1}^{N}\left\{2\pi\, \sigma^2 (\mu_i^0)^2\right\}^{-(r_i-1)/2} r_i^{-1/2} \exp\left\{\frac{-r_i\, s_i^2}{2\sigma^2(\mu_i^0)^2}\right\} .$$

Now from equation (10) (with V_1 replaced by σ^2) we can show that $\partial\mu_i^0/\partial\sigma^2 = -\mu_i^0/(2\sigma^2 + \zeta_i)$, where $\zeta_i = m_i/\mu_i^0$. Hence $\partial \log \Phi^0/\partial\sigma^2$ becomes

$$0 = -\frac{1}{\sigma^2}\sum_{i=1}^{N}(r_i - 1) + 2\sum_{i=1}^{N}\frac{(r_i-1)}{(2\sigma^2 + \zeta_i)}$$
$$+ \sum_{i=1}^{N}\frac{r}{\sigma^4}(\sigma^2 + \zeta_i - \zeta_i^2) - 2\sum_{i=1}^{N}\frac{r_i(\sigma^2 + \zeta_i - \zeta_i^2)}{\sigma^2(2\sigma^2 + \zeta_i)} . \qquad (20)$$

Thus the partial conditional maximum likelihood estimator of σ^2 is the solution V_3 of (20), where ζ_i at the solution is given by

$$\zeta_i = m_i/\mu_i^0 = (A_i/2) + \sqrt{((A_i^2/4) + A_i\, V_3)}$$

where

$$A_i = m_i^2/(s_i^2 + m_i^2), \text{ as before .}$$

When all the r_i's are equal we can investigate the consistency of V_3, as $N \to \infty$ by evaluating the expectation of the terms which are summed in equation (20). We consider the 'limit' equation

$$0 = \frac{(1-r)}{\psi} + 2(r-1)E\left(\frac{1}{2\psi+\zeta}\right) + rE\left(\frac{\psi+\zeta-\zeta^2}{\psi^2}\right) - 2rE\left\{\frac{\psi+\zeta-\zeta^2}{\psi(2\psi+\zeta)}\right\} . \qquad (21)$$

where we have dropped the suffices from A and ζ, as their distributions are independent of μ_i. Again, it is easy to see that this equation is not satisfied when $\psi = \sigma^2$, and so Jewell's partial conditional likelihood estimator is not consistent.

To examine the extent of this inconsistency for small σ^2 we can use the methods of previous sections for the expectations in (21). We seek a root of (21) of the form $\psi = a\sigma^2 + b\sigma^4 + \ldots$. We substitute this expression for ψ into (21) and, after some complicated evaluations of the expectations, equation (21) becomes

$$0 = \frac{(1-r)}{\psi} + (r-1)\frac{\sigma^2}{\psi^2} + 2(r-1) + \frac{\sigma^4}{\psi^2}\frac{(r-1)(1-2r)}{r} + O(\sigma^6) .$$

Equating coefficients in powers of σ^2 yields the root of (21) as

$$\psi = \sigma^2 + (1/r)\sigma^4 + O(\sigma^6). \qquad (22)$$

As in the last section (20) has multiple roots in general. Again it can be seen that (21) has a unique positive root ψ given by (22) for small σ^2. This root is the probability limit of roots of the equation (20) as $N \to \infty$ and it can be shown that equation (20) has at most one positive root, corresponding to the global maxima of the partial conditional likelihood. Thus, as before, the partial conditional likelihood estimator converges in probability to the root ψ.

In summary, for small σ^2, V_3 converges in probability to $\psi = \sigma^2 + (1/r)\sigma^4 + O(\sigma^6)$ and hence is identical to Raab's estimate V_2 to order σ^6. As in previous sections we can extend this conclusion to cover the case where the r_i's are not all equal but are consistently mixed as $N \to \infty$.

The behavior of ψ, the solution of (21), as $\sigma^2 \to \infty$ is similar to that of the maximum and modified likelihoods. The probability limit of V_3 as $N \to \infty$ tends to a finite number $\psi(r)$ as $\sigma^2 \to \infty$ and thus V_3 is bounded with probability one as $\sigma^2 \to \infty$. $\psi(r)$ is given by the root of the following equation in ψ.

$$-\frac{2(r-1)\,\Gamma(r/2)}{\sqrt{\pi}\,\Gamma((r-1)/2)}\int_0^1 \frac{(1-z)^{(r-3)/2}\,dz}{2\psi(z+4\psi)^{1/2}} + \frac{(r-1)}{(r+2)}\frac{1}{2\psi^2}$$

$$+\frac{r\,\Gamma(r/2)}{\sqrt{\pi}\,\Gamma((r-1)/2)}\int_0^1 \frac{(1-z)^{(r-1)/2}}{2\psi^2}\frac{(z+2\psi)}{(z+4\psi)^{(1/2)}}\,dz = 0 .$$

The values of $\psi(r)$ for $r = 2, 3$ are given by:

r =	$\psi(r)$
2	1.6961
3	3.6821 .

The partial conditional likelihood Φ contains some information on the nuisance parameters μ_i. If σ^2 is known we can maximize Φ with respect to the μ_i and obtain the estimators $\hat{\mu}_i = \left[r_i\, s_i^2/\{(r_i-1)\sigma^2\}\right]^{1/2}$. Now $r_i\, s_i^2$ is distributed as $\mu_i^2\,\sigma^2\,\chi^2_{(r_i-1)}$, so that the distributions, and in particular the variances, of the $\hat{\mu}_i$ are independent of σ^2. By contrast the estimator of μ_i from the partially sufficient statistic T_i is m_i which has variance $\sigma^2\,\mu_i^2/r_i$. Thus when σ^2 is small almost all the information on the μ_i is contained in the T_i, whereas when σ^2 is large most of the information on the μ_i is contained in Φ. Thus the partial conditional likelihood estimator in this example approaches consistency when Φ is approximately a true conditional likelihood.

Finally we note that there is another natural estimator which can be derived from Φ which behaves very differently, however, in this example, from Jewell's estimator. If we replace the μ_i in Φ, not by μ_i^0, but by the estimator m_i from T_i we obtain

$$V_4 = \sum_{i=1}^{N}(r_i\, s_i^2/m_i^2)/\sum_{i=1}^{N}(r_i-1) , \qquad (23)$$

i.e., an obvious estimator for σ^2 for any fully specified function $V(\mu)$. Now $r_i\, s_i^2/m_i^2$ is distributed as $(\sigma^2 Y/X^2)$ where Y is $\chi^2_{(r-1)}$ and independently X is $N(1,\sigma^2/r)$. Now we cannot use our previous methods to examine the probability limit of V_4 because Y/X^2 has an undefined expectation even for small σ^2. In fact when all the r_i's are equal V_4 does not converge to any limit in probability as $N \to \infty$. A proof is given in the Appendix. Further examination of this estimator shows that for any $\sigma^2 > 0$ V_4 has a distribution whose dispersion increases with N. Clearly V_4 does not suffer from the 'boundedness problem' of our other likelihood estimates as $\sigma^2 \to \infty$.

DISCUSSION

For the particular example $V = \mu^2$ we can restructure the problem to obtain a consistent estimator of the variance parameter σ^2. However this method cannot be extended to other variance functions. The methods which can be extended to other variance functions (modified likelihood, integrated likelihood and partial conditional likelihood) only produce approximately consistent estimators when the standard deviation of the observations is small compared with their mean value. Of these three methods only the partial conditional

likelihood approach can be extended to the general problem when we cannot obtain a non-constant conditional distribution independent of the nuisance parameters. It seems likely to be a useful technique when ϕ approximates a conditional likelihood. For the case of general V it remains an open question whether we can find a consistent estimator, as in the case $V = \mu^2$, or whether we must agree with Edwards (1972, p.109) that "in some cases the conditions under which we argue may have to include specific values for nuisance parameters."

We have nowhere discussed the efficiency of our estimation procedures. This is now being investigated by the authors, along with an extension of the methods discussed here to the case $V = |\mu|^{\theta}$.

APPENDIX

Proposition: There do not exist constants α_N such that $V_4^{(N)} - \alpha_N \xrightarrow{P} 0$ as $N \to \infty$ for any $\sigma^2 > 0$ where $V_4^{(N)}$ is given by (23) with $r_i = r$ $(i=1,\ldots,N)$.

Proof: For any $t > 0$,

$\Pr(V_4^{(N)} > t) = \Pr((r/r-1)(X/W > t))$ where X is distributed as $\chi^2_{(r-1)}$ and W is distributed as a non-central $\chi_1^2(\lambda)$ with non-centrality parameter $\lambda = r/\sigma^2$, independently of X.

$\therefore \quad \Pr(V_4^{(N)} > t) = \Pr(W < sX)$ where $s = (r/r-1)(1/t)$.

Now $\quad \Pr(W < s) = \sum_{j=0}^{\infty} ((e^{-\lambda}\lambda^j)/j!)\Pr(\chi^2_{(2j+1)} < s)$ since W is a Poisson mixture of central χ^2 variates (see Johnson and Kotz, 1970, p. 132). But

$$\Pr(\chi^2_{(2j+1)} < s) = \int_0^s (1/2)^{(2j+1)/2} \frac{x^{(2j-1)/2}e^{-x/2}}{\Gamma((2j+1)/2)}\,dx$$

$$> \frac{2^{-(2j+1)/2}e^{-s/2}s^{j+(1/2)}}{\Gamma(j+(1/2))(j+(1/2))} .$$

$$\therefore \quad \Pr(W < s) > \sum_{j=0}^{\infty} ((e^{-\lambda}\lambda^j)/j!)2^{-(j+(1/2))} \frac{e^{-s/2}s^{j+(1/2)}}{\Gamma(j+(1/2))(j+(1/2))} .$$

Now $\dfrac{1}{j!\,\Gamma(j+(1/2))(j+(1/2))} = \dfrac{2^{2j+1}}{\sqrt{\pi}\,(2j+1)!}$.

$$\therefore\quad \Pr(W < s) > \sum_{j=0}^{\infty} \left(e^{-\lambda}\lambda^{j}/(2j+1)!\right)\left(2^{j+(1/2)}/\sqrt{\pi}\right)\, e^{-s/2}\, s^{j+(1/2)}$$

$$= (e^{-\lambda}/\sqrt{\lambda\pi}) \sum_{j=0}^{\infty} \left((\sqrt{\lambda})^{2j+1}/(2j+1)!\right)\, (\sqrt{2})^{2j+1}\, e^{-s/2}(\sqrt{s})^{2j+1}$$

$$= (e^{-\lambda}/\sqrt{\lambda\pi})e^{-s/2}\, \sinh(\sqrt{2\lambda s})\ .$$

Hence
$$\Pr(W < sX) > \int_0^{\infty} (e^{-\lambda}/\sqrt{\lambda\pi})e^{-sx/2}\, \sinh(\sqrt{2\lambda s x})\frac{x^{(r-3)/2}e^{-x/2}}{2^{(r-1)/2}\Gamma((r-1)/2)}\, dx$$

$$= \left(c(\lambda)e^{\lambda s/s+1}/(s+1)^{(r-1)/2}\right)\left[\int_{-\sqrt{\frac{\lambda s}{s+1}}}^{\infty} e^{-V^2}\left(V+\sqrt{\frac{\lambda s}{s+1}}\right)^{r-2} dv - \int_{\sqrt{\frac{\lambda s}{s+1}}}^{\infty} e^{-V^2}\left(V-\sqrt{\frac{\lambda s}{s+1}}\right)^{r-2} dv\right]$$

where $c(\lambda)$ is a function of λ only. (*)

Now $t\Pr(V_4^{(N)} > t) = (r/r-1)(1/s)\Pr(W < sX)$ and thus from (*) we can easily see that $t\Pr(V_4^{(N)} > t) \to \infty$ as $t \to \infty$. Hence by a theorem of Kolmogorov (see Rao, 1973) the result follows.

REFERENCES

ANDERSEN, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Jour. Royal Stat. Soc.* B, *32*, 283-301.

BARTLETT, M.S. (1937). Properties of sufficiency and statistical tests. *Proc. Royal Soc.* A, *160*, 268-288.

BOX, G.E. & TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Mass.: Addison-Wesley.

COX, D.R. (1973). A note on partially Bayes inference in the linear model. *Biometrika*, *62*, 651-654.

DAWID, A.P., STONE, M. & ZIDEK, J.V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *J.R.Stat. Soc.*,B,35,189-233.

EDWARDS, A.W.F. (1972). *Likelihood*. Cambridge: Cambridge University Press.

FINNEY, D.J. & PHILLIPS, P. (1977). The form and estimation of a variance function, with particular reference to radioimmunoassay. *Appl. Stat.*, *26*, 312-320.

HEALY, M.J.R. (1979). Outliers in clinical chemistry quality control schemes. *Clin. Chem.*, *25*, 675-677.

JEWELL, N.P. (1979). Straight line fitting with heteroscedastic errors in both variables. (Preprint).

JOHNSON, N.L. & KOTZ, S. (1970). *Distributions in Statistics: Continous Univariate Distributions*, *Vol. 2*. New York: Wiley.

KALBFLEISCH, J.D. & SPROTT, D.A. (1970). Applications of likelihood methods to models involving large numbers of parameters (with discussion). *Jour. Royal Stat. Soc.*, B, *32*, 175-208.

KENDALL, M.G. & STEWART, A. (1969). *The Advanced Theory of Statistics, Vol. 1* (3rd Ed.). London: Griffin.

KIEFER, J. & WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, *27*, 887-906.

NEYMANN, J. & SCOTT, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, *16*, 1-32.

RAAB, G.M. (1979). Estimation of a variance function, with application to immunoassay. (To appear: *Applied Statistics*.)

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.

RODBARD, D., LENNOX, R.H., WRAY, H.L. & RAMSETH, D. (1976). Statistical characterization of the random errors in the radioimmunoassay dose-response variable. *Clin. Chem.*, *22*, 350-358.

SPROTT, D.A. (1973). Marginal and conditional sufficiency. *Biometrika*, *62*, 599-606.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Tech. Report #159, Series 2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONSISTENT ESTIMATION OF VARIANCE PARAMETERS FROM MANY SMALL SAMPLES WITH DIFFERENT MEANS		5. TYPE OF REPORT & PERIOD COVERED Tech. Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Nicholas P. Jewell and Gillian M. Raab		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0322
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N. J. 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 436) Arlington, Virginia 22217		12. REPORT DATE October 1979
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

79 12 14 079

18. SUPPLEMENTARY NOTES

Nicholas Jewell is affiliated with the Dept. of Statistics at Princeton University.
Gillian Raab is a member of the Medical Computing & Stat. Unit of the Univ. of Edinburgh, Edinburgh, Scotland

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. Abstract:

Where sets of observations are normally distributed with variances related to the mean of each set, the mean values become nuisance parameters when we wish to pool information about the variances from a large number of sets of information. This paper considers the problem of obtaining consistent estimators of the variance parameters where no assumptions or prior knowledge are available about the mean values. When the variance is proportional to the square of the mean we obtain an estimator for the constant of proportionality which is always consistent, by a marginal likelihood approach, but this method cannot be generalized to other variance functions. Integrated, modified, & partial conditional likelihood methods are investigated for this example and suggest methods for the general example when the standard deviation is small compared with the mean. The partial conditional likelihood method may be a useful general method of eliminating nuisance parameters in problems where the methods proposed by Kalbfleisch and Sprott (1970) are not applicable.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)